

# Investigation of a new multiphase flow choke correlation by linear and non-linear optimization methods and Monte Carlo sampling

Mohammad Babaei Khorzoughi · Mahmoud Safar Beiranvand ·  
Mohammad Reza Rasaei

Received: 7 March 2013 / Accepted: 1 July 2013 / Published online: 18 July 2013  
© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** In the production engineering point of view, the multiphase flow predictions from producing wells becomes important for an adept production design and management. As the associated costs of direct flow metering are considerable, the empirical and semi-empirical correlations are roughly efficient for this aim. Lack of sufficient data and their inherent epistemic and aleatory uncertainty due to human and equipment errors are, however, major limitations for these correlations. Following our previous article, in this paper, two optimization methods one linear and other non-linear are proposed. As it is shown, the linear regression is not dimensionally suitable to predict flow rate. It seems that due to the complexity of the objective function and also uncertain parameters, the linear regression is not the best algorithm for optimization. However, the non-linear method of Nelder–Mead (by means of MATLAB function *Fminsearch*) perfectly optimized the fitness function with a negligible average error. Due to the uncertain nature of main parameters in the correlation (such as  $P_{wh}$ , BS&W,  $T$ , etc.), a Monte Carlo sampling is used replacing these parameters with their PDFs (probability density function) to see if the proposed correlation works well or not. On this base, wellhead pressure ( $P_{wh}$ ), choke size ( $S$ ), basic sediment and water term ( $1 - \frac{BS\&W}{100}$ ), temperature ( $\frac{T}{T_{sc}}$ )

and gas/liquid ratio (GLR) are considered as random variables. The best probability distribution function (PDF) for each variable is then obtained which most closely reproduce flow through the choke. Monte Carlo sampling which deals with uncertain variables is used to predict the flow rates based on the proposed method and to show the level of uncertainty within the developed correlation results.

**Keywords** Monte Carlo simulation · Optimization methods · Basic sediment and water · Multiphase flow · Pearson correlation

## List of symbols

$A$	Proportionality constant
$B$	Bean or choke size exponent
BS&W	Basic sediment and water
$C$	Gas–oil ratio exponent
$D$	Basic sediment and water term exponent
$E$	Temperature term exponent
$F$	Wellhead pressure exponent
GLR	Producing gas/liquid ratio at standard conditions, scf/stb
GOR	Producing gas/oil ratio at standard conditions, scf/stb
OWC	Oil–water contact
$P_{wh}$	Wellhead pressure (psig)
$Q$	Gross liquid flow rate (bbl/day)
$Q_{correl}$	Calculated gross liquid flow rate from correlation (bbl/day)
$Q_{test}$	Measured gross liquid flow rate (bbl/day)
$S$	Choke or bean size, 1/64 in.
$T$	Temperature, $R$
$T_{sc}$	Standard condition temperature, $R$

M. Babaei Khorzoughi (✉)  
Department of Mining Engineering, University of British  
Columbia, Vancouver, Canada  
e-mail: mbabaei20@yahoo.ca

M. Safar Beiranvand · M. R. Rasaei  
Institute of Petroleum Engineering, School of Chemical  
Engineering, College of Engineering, University of Tehran,  
Tehran, Iran

## Introduction

The incorrect two- or three-phase flow rate estimations may cause optimum production failure and damage surface production facilities. In addition, in the field oil production period, it seems to be vital to predict the oilfield output for an efficient production management system (Guo and Deng 2009); and increasing the well's lifetime. Many multiphase choke correlations have been proposed by some scholars and researchers such as Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961) and Mesallati et al. (2000) between liquid flow rate, gas/liquid ratio, wellhead pressure and choke size. General form of these correlations and their coefficients are listed in Table 1. However, due to the highly complicated nature of the flow mechanisms and regimes in the pipes, wellheads and orifices, no comprehensive theoretical relation has been accepted under all operating conditions.

Uncertainty and risk analysis greatly help to explore the nature of complex systems such as choke behavior.

Basically, engineering judgment is needed when the lack of historical 'hard data' becomes an important issue (Apeland et al. 2002). Some researchers such as Aven and Pörn (1998) stated that quantitative risk assessment distinguishes between epistemic and aleatory (stochastic) uncertainties (Vaidogas and Juocevicius 2009). However, in the risk analysis, regarding the predictive and epistemic path, probabilities are used for expressing uncertainties related to future amounts of observable quantities in a time period (Apeland et al. 2002).

In the real world especially in the oil and gas industry, there are different kinds of these uncertainties associated with the measured or calculated parameters. To cope with these data uncertainties, several statistical approaches are developed. These approaches are the study of gathering, classification, analysis and interpretation of data. Some of the well-known statistical approaches are Monte Carlo Sampling, Goodness of fit tests, Hypothesis tests, etc.

In parametric statistics, linear regression is a mathematical method in which a straight line is fitted between a numbers of points to measure the explanatory variables effect on a scalar dependent variable (Hu 2011).

Monte Carlo simulation in production engineering is more or less new and is shown versatile for performance

prediction of individual wells during a field development plan. Considering the computational cost and time, Monte Carlo simulation is very useful to predict the accuracy of various estimation methods and their optimization in a wide range of applications (Ambrozic 2011).

Statistical analysis is used in this paper for parameter study of a new multiphase choke correlation that proposed (Safar Beiranvand and Babaei Khorzoughi 2012) recently. Interrelationship between some of its independent and dependent variables is then investigated. Getting access to a rich data bank, distribution function of each parameter of the new correlation is obtained. Monte Carlo analysis is then performed to fully explore the range of flow rate uncertainty that will be obtained from the proposed correlation at various operating conditions.

## Proposed correlation

In the previous paper (SPE-158649-PA), 182 production test data of different Iranian wells were used to propose the following correlation (Safar Beiranvand and Babaei Khorzoughi 2012):

$$Q = A \frac{P_{wh}^F S^B \left(1 - \frac{BS\&W}{100}\right)^D \left(\frac{T}{T_{sc}}\right)^E}{GLR^C} \quad (1)$$

where  $P_{wh}$  is wellhead pressure (psig), GLR is gas/liquid ratio (scf/stb),  $Q$  is gross liquid flow rate (bbl/day),  $S$  is choke size (64th of an inch),  $T$  is temperature (R) and BS&W is volume percentage of basic sediment and water in the producing fluids. These parameters ranges are given in Table 2. The field tests data used for this correlation are given in SPE-158649-PA.

It is shown in the previous paper that inclusion of  $\left(1 - \frac{BS\&W}{100}\right)$  and  $\left(\frac{T}{T_{sc}}\right)$  greatly improves the liquid flow rate prediction of the proposed correlation compare to other existing models. However, to explore the relation between these two parameters and liquid flow rate, a calculation of the Pearson product-moment correlation coefficient is done. It is obtained by dividing the covariance of the two variables by the product of their standard deviations. To conduct correlation analysis, Wessa (2013) statistics software is employed.

The Pearson product-moment correlation coefficient is obtained by the following relation (Rodgers and Nicewander 1988) and takes real values between +1 and −1 which correspond to the best positive and negative correlations, respectively (Dowdy and Wearden 1983):

$$P_{xy} = \frac{COV(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (2)$$

where COV is covariance and  $\sigma$  is standard deviation.

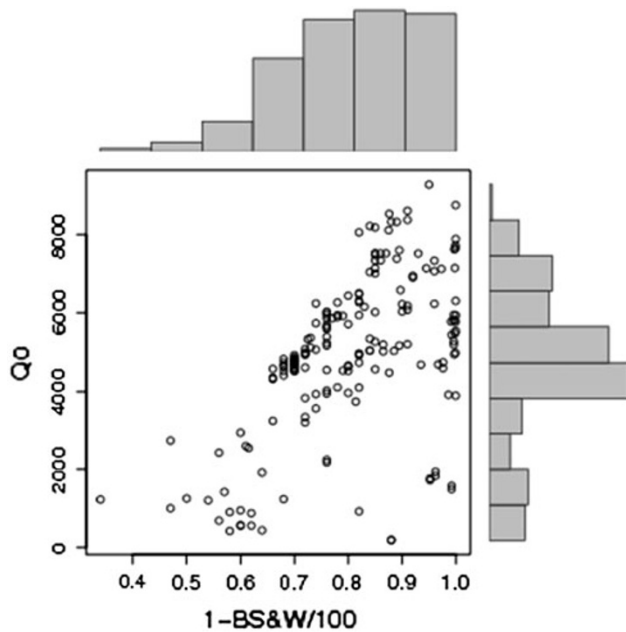
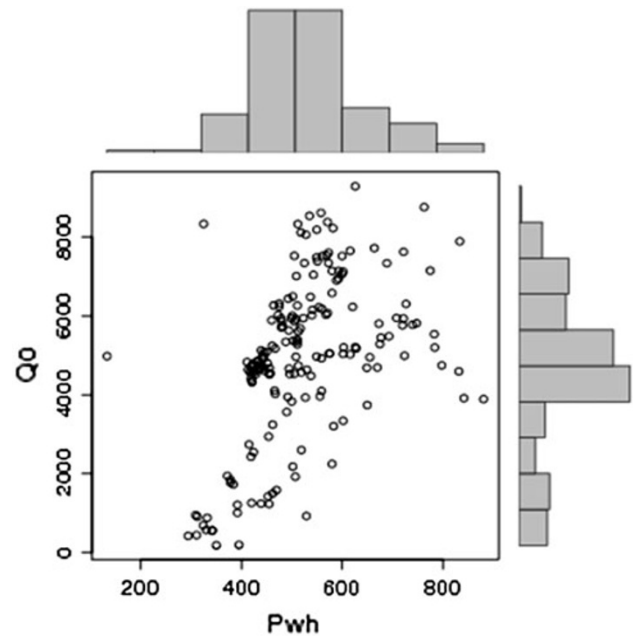
**Table 1** Multiphase flow choke correlations:  $Q = A \frac{P_{wh} S^B}{GLR^C}$

Correlation	A	B	C
Gilbert (1954)	0.1	1.89	0.546
Baxendell (1958)	0.104	1.93	0.546
Ros (1960)	0.574	2.0	0.5
Achong (1961)	0.2618	1.88	0.65
Mesallati et al. (2000)	0.0564	1.43	0.946



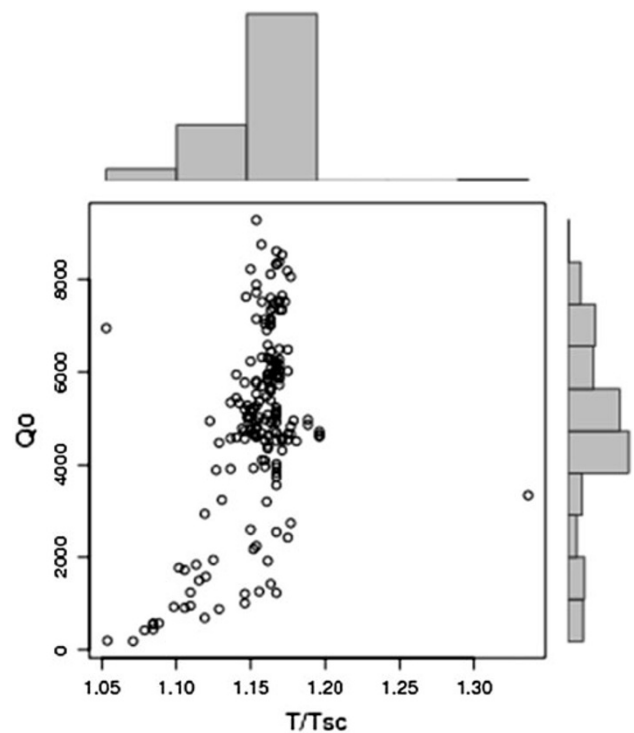
**Table 2** Range of data used for correlation

	Flow rate (stb/day)	GLR (scf/stb)	Choke size (1/64)in	Wellhead pressure (psig)	Temperature (°F)	BS&W (%)
Range	183–9284	36–885	25.6–40	133–883	87.6–162	0.1–53

**Fig. 1** Pearson product-moment correlation scatter plot of  $(1 - \frac{BS\&W}{100})$  vs.  $Q$ **Fig. 2** Pearson product-moment correlation scatter plot of  $P_{wh}$  vs.  $Q$ 

Figures 1, 2, 3 present the scatter plot of correlation coefficient between  $(1 - \frac{BS\&W}{100})$ ,  $(\frac{T}{T_{sc}})$  and  $P_{wh}$  versus  $Q$ , respectively. As depicted in the mentioned figures  $(1 - \frac{BS\&W}{100})$  and  $P_{wh}$  have stronger correlation with  $Q$  than  $\frac{T}{T_{sc}}$  and  $Q$ . In addition to the scatter plot, these figures show the histogram distribution of each parameter. So, the graph of  $Q$  versus  $(1 - \frac{BS\&W}{100})$  and  $P_{wh}$  are ascendant that is compatible with the physics of problem. In other hand, for  $Q$  versus  $\frac{T}{T_{sc}}$  it can be seen that for  $Q < 4,000$  bbl the relationship is ascendant but for  $Q > 4,000$  bbl, the  $\frac{T}{T_{sc}}$  is almost constant. Therefore, using the  $\frac{T}{T_{sc}}$  strongly depends on the other parameters.

All of mentioned parameters show moderate correlation coefficient in the range of 0.3–0.7 with flow rate which are shown in the Table 3. It is therefore evident that the new parameters of  $(1 - \frac{BS\&W}{100})$  and  $(\frac{T}{T_{sc}})$  versus  $Q$  have comparable correlation coefficient to that of between  $P_{wh}$  and  $Q$ . This analysis shows that these parameters have almost the same importance level in predicting  $Q$  and we should expect more prediction errors in previously published choke flow correlations compare to our proposed

**Fig. 3** Pearson product-moment correlation scatter plot of  $(\frac{T}{T_{sc}})$  vs.  $Q$

**Table 3** Correlation coefficient between  $Q$  and  $(1 - \frac{BS\&W}{100})$ ,  $P_{wh}$  and  $\frac{T}{T_{sc}}$ 

Variables	Correlation coefficient with $Q$
1-BS&W/100	0.543
$P_{wh}$	0.488
$\frac{T}{T_{sc}}$	0.462

correlation as will be explored shortly. For example, the correlation between  $Q$  and  $(1 - \frac{BS\&W}{100})$  is even stronger than correlation between  $Q$  and  $P_{wh}$ .

In addition to a correlation functional form and its variables, correlation parameters ( $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$  in this case) also greatly control its performance. As shown in Table 1, all the previous correlations proposed by other researchers share the same functional form and variables, but they have different parameters. It is therefore of great importance to correctly obtain these parameters. Two optimization methods for this purpose are used which will be explained in the next section.

### Determination of correlation parameters

Two optimization methods, linear regression and Nelder–Mead, are used to generate the unknown parameters  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$  of the proposed choke correlation.

**Table 4** Arbitrary initial values for  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$  and  $F$ 

Model parameters	Assigned bounds
$A$	[0 2]
$F$	[0.1 1]
$B$	[1 2]
$D$	[0.1 1]
$E$	[0.1 1]
$C$	[0.1 1]

**Table 5** Equation coefficient for different correlation

Correlation	Empirical coefficient						Error %	Absolute error %
	$A$	$B$	$F$	$C$	$D$	$E$		
New correlation	1	1.5	0.5	0.1	1	−0.8	2.89	31.6
Linear regression	0.0178	1.94	0.56	−0.047	0.73	6.82	N/A	N/A
Gilbert	0.1	1.89	1	0.546	0	0	60	67.6
Ros	0.574	2	1	0.5	0	0	160	160
Baxendell	0.1046	1.93	1	0.546	0	0	93.28	96.3
Achong	0.2618	1.88	1	0.65	0	0	137	139
Mesallati et al.	0.056	1.43	1.678	0.946	0	0	36.1	83.55

Performance goodness of these methods is quantified by calculating two types of errors as follows:

$$\text{error} = \frac{Q_{\text{test}} - Q_{\text{correl}}}{Q_{\text{test}}} \quad (3)$$

$$\text{absolute error} = \left| \frac{Q_{\text{test}} - Q_{\text{correl}}}{Q_{\text{test}}} \right| \quad (4)$$

### Nelder–Mead algorithm

In this approach the bounds based on the expert knowledge and previous models are applied into the model parameters, i.e.,  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$  in the present study. Based on the correlation (1) and the proposed bounds at each iteration (Table 4), the calculated values are compared with the actual values and the error is computed until an acceptable answer is reached. A script is written in MATLAB to perform the Nelder–Mead algorithm by means of MATLAB function `Fminsearch` based on arbitrary initial values for  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$  and  $F$ . The results are tabulated in the first row of Table 5. Small error of 2.89 % of the proposed correlation indicates close prediction of flow rates to the measured values compare to very unsatisfactory results of the previous correlations with errors in the range of 60–160 %.

### Linear regression

In this study, the general linear regression form is introduced by:

$$y = c_1x_1 + c_2x_2 + \dots + c_nx_n + \varepsilon \quad (5)$$

where  $y$  is dependent variable,  $c_i$  ( $i = 1, 2, 3, \dots, n$ ) is model parameter,  $x_i$  is independent variable and  $\varepsilon$  is error.

Rewriting correlation (1) in logarithmic form leads to the following relation more suitable to apply linear regression:

$$\ln(Q) = \ln(A) + F \ln(P_{wh}) + B \ln(S) + D \ln\left(1 - \frac{BS\&W}{100}\right) + E \ln\left(\frac{T}{T_{sc}}\right) - C \ln(\text{GLR}) \quad (6)$$

**Table 6** Probability distributions for random variables

Random variable	Probability distribution	Parameters
$P_{wh}$	Logistic	Mean = 514.79, scale = 65.24
$S$	Uniform	Min = 25.39, max = 64.21
1-BS&W/100	Beta	Min = -0.08, max = 1.17, alpha = 12.2557
$\frac{T}{T_{sc}}$	Student's $t$	Midpoint = 1.15, scale = 0.01, DOF = 2.99
GLR	Gamma	Location = 30.74, scale = 144.24, shape = 1.53
$Q$ (actual)	Weibull	Location = -7.7980.08, scale = 13,579.07, shape = 7.3

Compared to general linear regression form, the  $\ln(A)$ ,  $B$ ,  $C$ ,  $D$ ,  $E$  and  $F$  are model parameters and  $\ln(P_{wh})$ ,  $\ln(1 - \frac{BS\&W}{100})$ ,  $\ln(S)$ ,  $\ln(\frac{T}{T_{sc}})$  and  $\ln(GLR)$  are physical measurable parameters (explanatory variables).

In this study, a script is written in MATLAB to perform the linear regression algorithm. The results are shown in Table 5. This table shows that dimensions are not suitable due to the negative value for  $C$ .

### Uncertainty analysis of the proposed correlation

#### Statistical analysis

Statistical analysis is used here to investigate the inherent data uncertainty effect on the liquid flow rate calculated from the proposed correlation.  $P_{wh}$ ,  $S$ ,  $(1 - \frac{BS\&W}{100})$ ,  $\frac{T}{T_{sc}}$ , and GLR are considered as independent random variables and  $Q$  as dependent variable. It is worth to emphasize that a random variable does not necessarily refer to an unknown parameter, but also to a parameter with some uncertainty as it is here.

A random variable can be completely introduced by its probability density function (PDF). It is therefore necessary to determine the best PDF of the aforementioned parameters. This is achieved by functional analysis of histogram of each variable. Selected PDFs best fitted to the recorded variables histograms are given in the Tables 6, 7. The assigned distributions are based on the Chi-square goodness of fit. By applying these probability distribution functions in the Monte Carlo simulation of the new model, the most probable value of choke flow rate can be determined as explained in the next section. Choke flow rate changes as a function of considered operating variables is of utmost importance in production engineering and management.

#### Monte Carlo simulation

Monte Carlo simulation is a common method widely used in the field of probability analysis and risk management. It deals with uncertain variables to explore their

**Table 7** Most probable values based on the assigned distributions

Random variables	Most probable value
$Q$	4,119.775
$P_{wh}$	521.922
$S$	57.5578
1-BS&W/100	0.805337
$\frac{T}{T_{sc}}$	1.15472
GLR	251.663

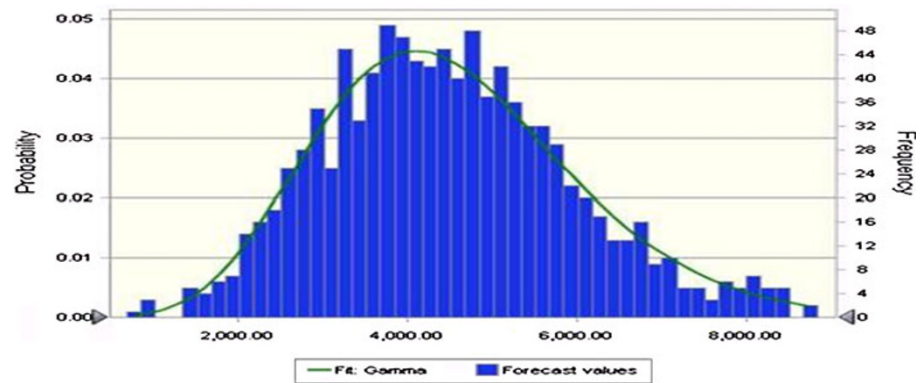
variation impact on an objective function in engineering or other mathematical and physical systems. In the Monte Carlo simulation uncertainties variables are modeled based on their statistical distribution and random number sampling. In the recent decades this approach has attracted researcher's attention to estimate uncertain parameters and to do risk analysis in oil and gas industry.

As previously discussed, one of the most important issues in production engineering is to predict flow behavior in multiphase systems. However, one of the challenges in predication of flow rate in the two-phase systems is that it is prone to error. This error includes instrumentation errors as well as human (operator) errors. All of the variables gathered from the field to predict flow rate include these errors. Therefore, a calibration or an index as a measure of accuracy is required to have more accurate results.

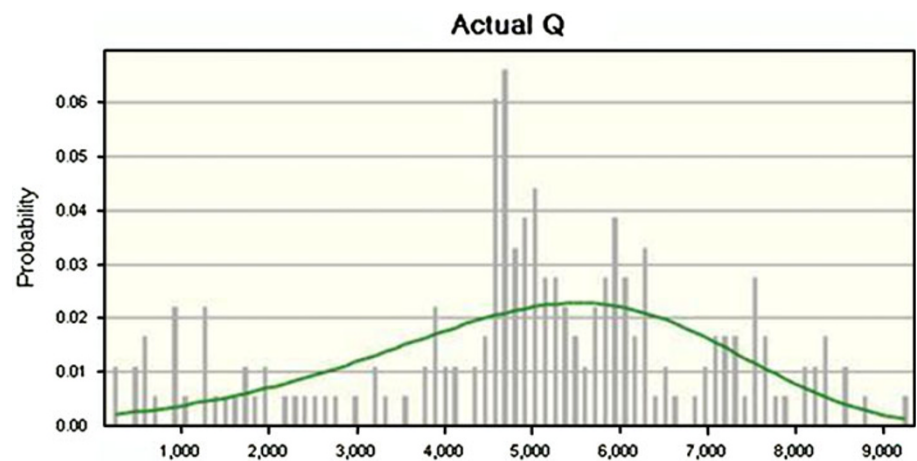
In the current investigation, Monte Carlo simulation is used to forecast the probability distribution function (PDF) of choke flow rate,  $Q$ . The comparison between measured flow rates in the field and simulated values shows the level of uncertainty. Sampling from the PDFs of the independent physical variables is performed to obtain chock flow rate from correlation (1). Total number of 1,000 trials are performed to cover the physical variables ranges and provide proper flow rate distribution. The results are shown in Table 7 and in Fig. 4. Gamma distribution is obtained for  $Q$  with the most probable value of 4,120 stb/day. This Monte Carlo simulated distribution compares very well to the actual chock flow rate distribution in Figs. 5, 4. Although the distributions are not exactly the same, but it is



**Fig. 4** Probability distribution of  $Q$  due Monte carlo simulation



**Fig. 5** Probability distribution of actual  $Q$  from field data



clear from this analysis that the proposed correlation can reproduce the actual choke flow rate distribution from the considered random variables in their physical recorded ranges. The difference between these probability density functions was caused by the uncertain variables.

### Summary and recommendations

It is shown that the empirical correlations of Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961) and Mesallati et al. (2000) failed to predict choke flow rate of 182 production test data of Iranian wells. Prediction errors of 60–160 % are obtained from these correlations. Two new physical variables of  $(1 - \frac{BS \& W}{100})$  and  $(\frac{T}{T_{sc}})$  are shown to be correlated with choke flow rate. Inclusion of these variables into a new correlation greatly improved its liquid flow rate prediction. Determination of coefficient parameters of the proposed correlation by Nelder–Mead method resulted in only 2.89 % prediction error of the production test data.

Monte Carlo simulation of the proposed correlation showed a gamma distribution for choke flow rate  $Q$ , with the most probable value of 4,120 stb/day. This was in very

good agreement with the actual choke flow rate distribution from the recorded test data. The proposed correlation can therefore be effectively used to predict choke flow rate in the case of lack of information or large uncertainty in the production recorded data.

Regression results shows that linear regression is not appropriate for this problem because the dimensions are not suitable due to negative value for  $C$  which disturbs the dimensional equilibrium. It shows that the optimization method should be selected due to physical properties of problem. Hence, Nelder–Mead method is applicable in some cases of production problems.

Based on the difference in the actual  $Q$  and forecasted distribution, the coefficient of parameters in the proposed equation should be improved to obtain more accurate results.

Other new optimization techniques, especially evolutionary algorithms, such as Genetic Algorithm, Artificial Ant Colony, Simulated Annealing and Imperialist competitive algorithm, are recommended for further researches in production engineering such as multiphase flow prediction through the surface choke.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use,

distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## Appendix

See Table 8

**Table 8** Pearson product-moment correlation for different variables vs.  $Q$

Statistic	Variable X	Variable Y
Pearson product moment correlation for variable for $(1 - \frac{BS&W}{10})$ vs. $Q$		
Mean	0.805	4932.852
Biased variance	0.017	4195834.203
Biased standard deviation	0.132	2048.373
Statistic		
Covariance	147.763	
Correlation	0.543	
Determination	0.295	
$T$ test	8.671	
$P$ value (2 sided)	2.442e–15	
$P$ value (1 sided)	1.221e–15	
Degrees of freedom	180	
Number of observations	182	
Pearson product moment correlation for variable $P_{wh}$ vs. $Q$		
Mean	521.922	4932.852
Biased variance	13993.373	4195834.203
Biased standard deviation	118.293	2048.373
Statistic		
Covariance	118906.972	
Correlation	0.488	
Determination	0.238	
$T$ test	7.501	
$P$ value (2 sided)	2.787E–12	
$P$ value (1 sided)	1.393E–12	
Degrees of freedom	180	
Number of observations	182	
Pearson product moment correlation for variable for $(\frac{T}{T_{sc}})$ vs. $Q$		
Mean	1.154	4932.852
Biased variance	0.00076	4195834.203
Biased standard deviation	0.0275	2048.373
Statistic		
Covariance	26.274	
Correlation	0.462	
Determination	0.214	
$T$ test	6.999	
$P$ value (2 sided)	4.918E–11	
$P$ value (1 sided)	2.459E–11	
Degrees of freedom	180	
Number of observations	182	

## References

- Achong I (1961) Revised bean performance formula for Lake Maracaibo wells. Internal report. Shell Oil Co., Houston
- Ambrozic Milan, Gorjan L (2011) Reliability of a Weibull analysis using the maximum-likelihood method. *J Mater Sci* 46:1862–1869
- Apeland S, Aven T, Nilsen T (2002) Quantifying uncertainty under a predictive, epistemic approach to risk analysis. *Reliab Eng Syst Saf* 75:93–102
- Aven T, Pörn K (1998) Expressing and interpreting the results of quantitative risk analyses. *Rev Discuss, Reliab Eng Syst Saf* 61(1):3–10
- Baxendell PB (1958) Producing wells on casing flow—an analysis of flowing pressure gradients. In: *Transaction of the American institute of mining, metallurgical, and petroleum Engineers*, vol 213. Society of Petroleum Engineers, Dallas, Texas, pp 202–207
- Dowdy S, Wearden S (1983) *Statistics for research*. Wiley. ISBN 0-471-08602-9, pp 230
- Gilbert WE (1954) Flowing and gas-lift well performance. In: *Drilling and production practice*, vol 20. American Petroleum Institute, pp 126–157
- Guo L, Deng X (2009) Application of improved multiple linear regression method in oilfield output forecasting. In: *Proceedings of the ICIII'09, 2009 International conference on information management, Innovation management and industrial engineering-Volume 01*. ISBN: 978-0-7695-3876-1
- Hu Y (2011) Linear regression 101. *J Valid Technol* 17(2):15. ISSN 1079-6630
- Mesallati A, Biznati M, Mansouri N (2000) Multiphase-flow choke correlations for offshore Bouri oil field. In: *Proceedings of international gas union 21st world gas conference*, Nice, France, June 6–9, p 123
- Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. *Am Stat* 42(1):59–66
- Ros NCJ (1960) An analysis of critical simultaneous gas/liquid flow through a restriction and its application to flow metering. *Appl Sci Res* 9(1):374–388
- Safar Beiranvand M, Babaei Khorzoughi M (2012) Introducing a new correlation for multiphase flow through surface chokes with newly incorporated parameters. *SPE Prod Oper* 27(4):422–428. doi:10.2118/158649-PA SPE-158649-PA
- Vaidogas ER, Juocevičius V (2009) Assessment of structures subjected to accidental actions using crisp and uncertain fragility functions. *J Civ Eng Manag* 15(1):95–104
- Wessa P (2013) Free statistics software, office for research development and education, version 1.1.23-r7, URL <http://www.wessa.net/>